CREATING STATISTICALLY LITERATE GLOBAL CITIZENS: THE USE OF INTEGRATED CENSUS MICRODATA IN TEACHING

Ann Meier[1], Robert McCaa[2], and David Lam[3]
[1]Department of Sociology and Minnesota Population Center, University of Minnesota
[2]Department of History and Minnesota Population Center, University of Minnesota
[3]Department of Economics and Population Studies Center, University of Michigan
United States of America
meierann@umn.edu

We describe the use of international census micro data for teaching undergraduate students in two population related courses. The Integrated Public Use Micro Series (IPUMS) of international census data contains social and population data from 84 countries from 1960 to the present. We detail the use of the IPUMS data in courses at the University of Minnesota and the University of Michigan to help students understand population issues in historical and comparative perspective. One of the most significant learning outcomes of these courses was the acquisition of a baseline level of statistical literacy among non-statistics students in the social sciences. An added benefit was the global understanding gained through the investigation of countries other than their own across time and in comparison to other countries.

CENSUS MICRODATA AND THE IPUMS-INTERNATIONAL PROJECT

Census microdata are the individual responses to census questionnaires recorded in computerized form as numeric or alphabetic codes. The data include demographic characteristics such as age, sex, marital status, education, and occupation among other individual-level variables. They also include information on household characteristics such as urban/rural living, home ownership, and access to utilities like electricity and water supply. Over the past half century most of the major statistical agencies have prepared census microdata files for analysis by staff. Analysis of large census microdata files is now possible for ordinary researchers and even students.

The idea of gaining access to census microdata from around the world and across time is exciting but daunting. The first challenge is gaining access to the data from the countries of interest. Access is improving, but it is still challenging to individual researchers. A second challenge is comparability. To trace trends across time in a country or to compare countries to one another, files from different years and countries must first be made compatible. Historically, there has been little coordination between national statistical agencies to facilitate comparison. Even the same statistical agency sometimes asks questions differently in different census years.

The Integrated Public Use Micro Series–International (IPUMS-I) project is a global initiative in cooperation with national statistical authorities world-wide to anonymize, integrate, and disseminate samples of census microdata to researchers, policy makers, teachers, and students. As of August 2009, IPUMS is the largest repository of census microdata in the world with the official statistical authorities of more than 84 countries, encompassing more than four-fifths of the world's population, entrusting a total of 240 censuses to the Minnesota Population Center.

The IPUMS has two goals: first, to preserve census microdata and, second, to make anonymized, integrated sample extracts available to researchers and policy analysts free of charge. This article focuses on the latter goal, and offers an illustrative example of how the IPUMS effort facilitates statistical literacy and general global awareness among the next generation.

To make census microdata useful for research, they must be thoroughly documented and integrated. The basic goal of the IPUMS harmonization efforts is to simplify the use of the data while losing no meaningful information. This is challenging because to make the data simple for comparison across time or place, it is necessary to create comparable codes across samples. Therefore, the harmonized measures must be the lowest common denominator from all the samples that include a measure of a particular characteristic. For example, regarding information on school attendance, all samples have enough information to indicate whether the respondent is currently in school or not. However, some samples contain further information indicating, for those who are not in school, whether they attended school in the past or never attended. To avoid the loss of important information, the IPUMS uses a composite coding strategy that retains all original detail,

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

and at the same time provides comparable codes across samples. The first one or two digits of each code provide information that is available across all samples. The next digits provide additional information available in a subset of the samples. The trailing digits provide detail that is only rarely available. Where information is not available for a sample, a zero place-holder is assigned to that digit. The harmonization allows researchers to use detailed measures on a single country or less detailed but comparable measures across countries, years, or countries and years. Thus, the IPUMS database has become a very flexible and user-friendly data source.

In addition to informing development strategies and policy decisions, IPUMS is now used as a teaching tool. Using IPUMS data in teaching serves several important pedagogical goals. First, it promotes statistical literacy among students who may learn statistical theory in the classroom, but rarely have the opportunity to apply their knowledge to real-world data. Second, it allows students to gain insight and knowledge about countries world-wide and across time. Below we offer two examples of how we have used IPUMS data in our undergraduate classrooms.

COURSE 1: WORLD POPULATION PROBLEMS (UNIVERSITY OF MINNESOTA)

Ann Meier used the IPUMS data in her World Population Problems course. The course is an upper level undergraduate course, enrolling mostly juniors and seniors. Many of the 55 students enrolled were American, but a few were foreign born, several were first-generation immigrants and at least ten students had parents who had immigrated to the U.S. before the students were born.. The students and instructor had a keen interest in expanding their world views.

Population Profile Assignments

We used a text book, and the instructor delivered a series of lectures on topics ranging from global population trends to family planning to population and economic development. The highlight of the students' learning, however, was a series of assignments in which students accessed and analyzed census data from the IPUMS project along with other sources of data. Each student was assigned a new identity: they were assigned a gender, age, and country of residence, along with a year in which they were living. For example, a female student was assigned an identity of a 22-year-old male from Iraq in 1997. We created mock "passports" to detail each student's new identity. The assigned countries and years corresponded with available IPUMS samples (e.g. Iraq census 1997), and three or four students were assigned the same country but different ages, genders, and census years. Students assumed their new identities when completing a series of assignments throughout the semester, culminating in poster session at the end of the semester. Using the IPUMS data and other sources in conjunction with new student identities was meant to help students understand how population issues are experienced differently based on geography (where you live), history (when you live), and social location (your age and gender).

In the first weeks of the semester, students learned about health and mortality. They accessed tables from the United Nations World Population Prospects to document the crude death rate, infant mortality rate, and life expectancy for their assigned country and year. To assess how different living conditions influence health, students used the IPUMS data to investigate infrastructure features such as source of water supply and access to flush toilets. Students used SPSS software to generate simple descriptive statistics using the IPUMS data. With their statistical output, they created a health and mortality profile for someone of their assigned identity. They combined statistics, time-trend and regional graphics, and a text evaluation of the health and mortality situation.

Next, the course turned to discussion of fertility and family situations. Again, students accessed data from IPUMS for their country and year to understand marital status, number of children, and families per household. In their final assignment, students joined with others who had been assigned identities of the same country to combine what they had about the country. The posters were displayed in a poster session that was widely attended. Students were evaluated on the content and presentation of their posters as well as how well they answered questions.

Learning Outcomes

The series of population profile assignments resulted in four key learning outcomes. First, students became experts on the population characteristics in a country other than their own. Being

assigned a new identity with a mock "passport" personalized the assignments, and therefore made students more invested in doing a good job. In addition, having to present what they learned and answer questions from a live audience motivated students to gain competency in. Second, students learned how history and geography shape population issues. By comparing health, mortality, fertility and family statistics across time and countries, students gained a better understanding of how historical events and geographic location can influence population. Third, students learned how to analyze secondary data by using the IPUMS database and SPSS software. This is a practical skill that they can carry forward into their other coursework and future professional lives. Many students expressed surprise and pride in their ability to easily generate graphs and tables based on their analysis of IPUMS census microdata. Finally, the availability of the IPUMS data helped make the world seem like a smaller place for students; it helped them feel more connected to those from other parts of the world. It is our hope that this makes them better global citizens.

COURSE 2: ECONOMICS OF POPULATION (UNIVERSITY OF MICHIGAN)

David Lam and his colleague Martha Bailey have incorporated IPUMS-I data into their course on Economics of Population. This is an upper-level undergraduate economics course. The course meets in a lecture format on Mondays and in a computer lab on Wednesdays. The class is designed to satisfy the requirement that students take an advanced writing course in their third or fourth year, ideally within their major field of concentration. The IPUMS-I data are directly linked with numerous writing assignments, including a major term paper.

Computer Labs with IPUMS-I

The course assumes that students have no prior experience with the statistical package Stata, although a few students typically have used it in prior courses. Students work with Stata in the early labs to learn the basics of the software and how census data are collected and distributed for public use. Weekly problem sets are used to reinforce material covered in the lab. For example, initial labs involve making simple tables of the distribution of children ever born by mother's age, generating new variables such as the natural logarithm of household income, and writing simple Stata programs. IPUMS-I is an excellent resource for teaching students a statistical package like Stata. Because variables have been integrated across years and countries it is easy, for example, to use the 2000 United States census during the instructional lab sessions, then have the students do a problem set that does similar analyses using the 1999 Kenya census. Once the students have been taught to append the data sets it is easy for them to do a problem set that compares two countries.

The course assumes that students have previously taken intermediate microeconomics and introductory courses in statistics. Therefore, the course moves quickly into multivariate regression analysis. One lab, for example, has students estimate regressions using number of children ever born as the dependent variable. The students use various combinations of the education of the mother, the education of the father, and household income as the independent variables, discussing how the results change as different variables are included. This gives students experience at interpreting regression coefficients and understanding the impact of including additional variables in a multivariate regression. Although the students usually have relatively good knowledge of statistics and econometrics, it takes students several weeks and many different examples before they are comfortable with interpreting regression coefficients in real-world examples.

One advantage of the IPUMS-I data is that the large sample sizes make it possible to include many variables in a regression, along with interactions and higher-order terms, while still producing meaningful estimates. The large sample sizes also make it possible to produce separate estimates for population subgroups such as narrowly defined age groups, regions, and racial and ethnic groups. A very attractive feature of the IPUMS-I data extraction system is that it is easy to extract samples with a target file size. For computer lab purposes we typically use extract files that are less than 50MB in size to keep computer processing fast. Students are encouraged to work with larger extracts for their projects, however, and it is easy for them to generate customized files that meet their needs. For example, one student wrote an excellent paper analyzing schooling outcomes for the children of Mexican immigrants in the state of California. He built an extract using only California residents in several recent U.S. censuses, giving him the statistical power necessary for his analyses while keeping the file size manageable.

Learning Outcomes

The course combines lectures, computer labs, problem sets and writing assignments in a systematic way, culminating in a 20-page term paper built around statistical analysis of IPUMS-I data. Students are encouraged to analyze changes in one country over time or to do a comparative analysis of two or more countries. Students write a short proposal in the fourth week of the course in which they take data from one country and present a few tables and figures analyzing some set of variables. They get feedback on these proposals, and the first draft is due halfway through the semester. They get extensive feedback on this draft, which is then revised before final submission. Considerable time is spent teaching students how to make tables and figures that effectively present their statistical results. They are told to write their papers as data-intensive policy briefs, focusing on how to communicate statistical analysis in a clear and effective way.

Students have shown a great deal of creativity in using IPUMS-I data. Examples of term papers include: Poverty and education in Kenya; Living arrangements of the elderly in South Africa; Women's education and fertility decline in Brazil; The impact of conflict on children's literacy in Rwanda; Male-female earnings gaps across time and across countries; The impact of family size on educational attainment in Costa Rica.

Student feedback from the course is very positive. Many students said the course allows them to apply the statistics they have learned previously to real-world economic problems. In the process they understand statistics better and appreciate the power of statistics to inform public policy. By the end of the course the best students are doing applied econometric analysis at a graduate level. Other students do not quite reach that level, but they learn how data can be used for statistical analysis and about the challenges in drawing causal inferences.

The skills acquired in the course often have immediate payoff. A number of the top students have taken prestigious research assistant positions in Washington agencies such as the World Bank, the Federal Reserve Board, and the Urban Institute. These jobs are standard stepping stones to the top doctoral programs in the country. Other students have secured jobs in government and private industry where their ability to carry out statistical analysis of large data sets is highly valued. All students develop a better understanding of how statistics can be applied using large-scale data to answer important questions. Many of them develop an appreciation for the enormous differences in living standards within and across countries.

CONCLUSIONS

In summary, the IPUMS International project offers researchers, teachers and their students access to the world's largest collection of census microdata. Every few months a new statistical agency signs on to this initiative (see https://international.ipums.org/international/). As demonstrated above, the IPUMS data provide a unique opportunity for students to learn new skills and gain important understanding about other countries. The skills contribute greatly to students' statistical literacy, while the understanding makes them more informed and compassionate global citizens. We encourage others to adapt these examples of the use of IPUMS data for their own teaching needs or develop their own creative use of the IPUMS data.

REFERENCES

Chase, C., Anekwe, T., Barofsky, J., & Farshad, F. (2008). "Economic effects of malaria eradication: evidence from an eradication experiment in Kigezi, Uganda". Working Paper.

Esteve, A., & Sobek, M. 2003. Challenges and Methods of International Census Harmonization. *Historical Methods, 36,* 66-79.

McCaa, R., Esteve, A., Ruggles, S., & Sobek, M. (2006). Using Integrated Census Microdata for Evidence-based Policy Making: the IPUMS-International Global Initiative. *The African Statistical Journal, 2,* 83-100.

Thomas, K. J. (2008). Return Migration in Africa and the Relationship between Educational Attainment and Labor Market Success: Evidence from Uganda. *International Migration Review, 42*(3), 652-674.

Ruggles, S., Sobek, M., et. al. 1997. *Integrated Public Use Microdata Series: Version 2.0*. Minneapolis: Historical Census Projects, University of Minnesota.